Control of High Performance Computing (HPC) tasks emissions

Context: Inria (Spirals team) and Qarnot CIFRE grant

The digital industry is booming, partly due to the intensive digitalization of most of our personal and professional activities. However, in view of the general awareness of the environmental impact of digital technology, many questions are being asked as to the validity of a race to develop new computing infrastructures that do not take this footprint into account.

In this field, the company Qarnot - https://qarnot.com/ - stands out from traditional infrastructures by proposing a decentralized cloud solution that allows the heat emitted by servers to be reused to power sites with high heat requirements: swimming pools, heating networks, industries, etc.

The Spirals team - https://team.inria.fr/spirals - is a joint project-team of Inria[1] and the University of Lille within the CRIStAL research laboratory. Spirals conducts research activities in the fields of distributed systems and software engineering. Spirals aims at introducing more automation in the adaptation mechanisms of software systems, especially around the issues of energy consumption.

In order to better understand the environmental impact of edge computing infrastructures, Qarnot, Inria and Ademe[2] have joined forces in a joint challenge, the four-year PULSE (PUshing Low-carbon Services towards the Edge) project. This project aims to measure, understand and optimize the energy consumption of distributed computing infrastructures such as the one proposed by Qarnot.

Objective

In tomorrow's computing services, users will have objectives in terms of quality of service and cost, to which will be added new objectives related to the energy impact and environmental footprint of their calculations. As these objectives are contradictory, there will necessarily be a compromise to ensure. This thesis will focus on the realization of this trade-off at the level of a computational task, or a cluster of computational tasks of a same user, the user deciding on the relative importance to be brought to the parameters mentioned above.

The management of the performance/cost/carbon trade-off must be done in an autonomic way, i.e., dynamically, during the execution of tasks, and according to the current state of the task [1]. Indeed, many variations can come from user objectives, task behavior, and variable infrastructure conditions generating constraints.

---

[1] French National Institute for Research in Digital Science and Technology
[2] The French Agency for Ecological Transition

Missions

To solve this problem, the PhD student will have to use the tools of automatic control to reinforce the orchestration methods of Qarnot. Control theory has indeed been used in the last few years to address computational management issues [2] and has made it possible to make progress on topics that are difficult to address with traditional tools in this field, especially those related to temporal constraints.

A first analysis of the action levers available in the medium term has allowed us to draw up a list of those to be considered, such as task preemption or virtual machine migration. The first step is to deepen these levers of action and to define the interactions between the trade-off mechanism - the controller, the current Qarnot orchestrator, and the infrastructures. In a second step, we will focus on the design of the dynamic decision mechanism, that is robust to variations in the user trade-off and to variations in the computational load.

Research-oriented aspects of this topic include:

- Review of the state of the art on the management of computational systems using automatic control, e.g. [3],
- Scientific contributions to the definition of the problem according to the formalism of control theory, the study of solutions and their evaluation,
- Dissemination of results through publications and participation in international top conferences.

Required skills

The candidate must hold an Engineering degree or a Master's degree in computer science or automatic control.
As this thesis lies at the border between computer science and control, expertise and technical knowledge are not expected on all fields, and therefore will differ according to the candidate's profile. This thesis will be an opportunity to develop skills on transversal subjects. Regardless the initial profile, an interest in the environmental impact of information technologies is however highly desirable.

Computer scientist profile:

- Programming languages: working knowledge of at least one modern object language (C#, Java, Python, etc.).
- Cloud, Virtualization, Docker
- Database (e.g. MariaDB, MongoDB)
- Development and collaboration tools (Git, Gitlab)

Control scientist profile:

- Definition of a control problem, frequency domain or state space representations
- Modeling, identification from data, estimation
- Controller design (optimal, adaptive, robust, non-linear, etc.)
- Analysis and simulation (Matlab, Python)
- Experimentation on real systems and data

Languages: English (French desirable)

Additional information

- Location: Paris and Lille (to be defined)
- Supervisors : Sophie Cerf (Inria Spirals), Lionel Seinturier (Inria Spirals), Alexandre Van Kempen (Qarnot)
- Type of contract: 36-month CIFRE fixed-term contract starting in September 2023

Application

The application file must include:

- a resume
- a cover letter
- transcripts of grades from the master/engineer program
- a recommendation letter (optional)

Application deadline: June 16, 2023

References

[1] Kephart, J. O., & Chess, D. M. (2003). The vision of autonomic computing. *Computer*, *36*(1), 41-50.

[2] Filieri, A., Maggio, M., Angelopoulos, K., d'Ippolito, N., Gerostathopoulos, I., Hempel, A. B., … & Vogel, T. (2015, May). Software engineering meets control theory. In 2015 IEEE/ACM 10th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (pp. 71-82). IEEE.

[3] Sophie Cerf, Raphaël Bleuse, Valentin Reis, Swann Perarnau, Eric Rutten. Sustaining Performance While Reducing Energy Consumption: A Control Theory Approach. EURO-PAR 2021 - 27th International European Conference on Parallel and Distributed Computing, Aug 2021, Lisbon, Portugal. pp.334-349, ⟨10.1007/978-3-030-85665-6_21⟩. ⟨hal-03259316⟩